Natural Hazards and Earth System Sciences Discussions

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC3
https://doi.org/10.5194/nhess-2022-58-RC3, 2022

Comment on nhess-2022-58

Anonymous Referee #2

Referee comment on "Improving the predictability of the Qendresa medicane by the assimilation of Conventional and Atmospheric Motion Vector observations. Storm-scale analysis and short-range forecast." by Diego Saúl Carrió, Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2022-58-RC3, 2022

Reviews of "Challenges assessing the effect of AMVs to improve the predictability of a medicane weather event using the EnKF. Storm-scale analysis and short-range forecast" (nhess-2022-58)

The author describes and implements an assimilation system to improve the prediction of medicane Qendresa of 2014 using atmospheric motion vectors (AMVs) and other observations. The author presents the result of four ensemble simulations: one without data assimilation and 3 experiments with data assimilation (SYN, RSAMV, CNTR, section 3.4).

1. The author mentions that not all assimilative ensemble members represent the medicane (roughly only the half of the ensemble includes the mediane; exact number of the experiments are on page 20). The manuscript did not mention how many of the ensemble members for the free ensemble model run includes the medicane but judging from figure 11, it seems more ensemble members of the free ensemble represent this medicane. Unfortunately, this is not addressed in the manuscript. It would seem to me a necessary first step, why apparently the assimilation inhibits the generation/developpement of the medicanes and to improve this aspect.

The model validation statistics later are only based on the subset of model members representing the medicane which can result in a misleading interpretation. For the sake of argument, let's take an extreme case when only one ensemble member would represent the medicane track but with good accuracy. The average track would then be simply equal to this medicane trak. In an ensemble would represent 11 members with an mediane, one with a good trak and one with a 10 a biased track, then the average will be worse than the first ensemble average track. In a forecasting scenario you are not sure whether actually a mediane will develop or not, the latter example (ensemble with 11 medicanes) seems to be more informative to me. However, by computing the average only over members including the medicanes favors in this case, the ensemble members with fewer medicanes.

To make this more concrete, we can use Figure 15 where the atmospheric pressure at the Malta airport is shown. In a forecasting scenario, your best estimate of the atmospheric pressure would be the ensemble mean using all ensemble members (including those who do not include the medicane) as you do not know yet if the medicane will actually develop or dissipate early. From Figure 15, it even seems that all 36 ensemble members of the free ensemble simulation were used (even those with a very weak depression) which would bias the results even more towards the assimilation simulation where the ensemble members without medicane are excluded.

2. Even when using their selective validation approach, the improvements from the assimilation are not so clear. For instance on Figure 13, it is not so clear to me that overall the assimilation experiments (SYN, RSAMV, CNTR) are actually better than ensemble with data assimilation given the large bias at the end. Is it also surprising that the type of the observations assimilation (which are very diverse) do not seem to matter much.

3. Another problem is that the manuscript does not show clearly which comparisons are against independent data and which are against assimilated data. In Figure 9, a correlation coefficient between model and maritime buoys of 0.996 is presented. If this is assimilated data, then you could easily have an correlation coefficient almost equal to 1, if you let the error variance of these observations tend to zero. But clearly, this would lead to a highly degraded forecast.

These major points are further developed below.

For the mentioned reasons I recommend major revision before publishing this paper.

More specific comments

==========

Line 55: "Although both methods are slightly different and contain different

types of errors associated, the overall information drawn from them has been found to be equivalent (Migliorini, 2012). From these reasons, in this study only satellite-derived products will be considered."

The cited study Migliorini (2012) does not show that the assimilation for both methods is an all cases equivalent. It rather defines two testable conditions, under which both approaches lead to equivalent results:

"(i) the radiance observation operator needs to be approximately linear in a region of the state space centered at the retrieval and with a radius of the order of the retrieval error; and (ii) any prior information used to constrain the retrieval should not underrepresent the variability of the state, so as to retain the information content of the measurements."

The author should check that two conditions deduced by Migliorini (2012) are verified before stating that both methods are equivalent.

Line 92: "Among the different available medicanes, the so called Qendresa, which took place southern Sicily between 7-8 November 2014 (Carrió et al., 2017) and was poorly forecasted,

was selected to perform this study. More precisely, the correct prediction of both the northward loop trajectory followed by Qendresa and its intensification still remain a major challenge for most current numerical weather models."

Can you be more specific which models did a poor forecast and show a figure of the forecast and actual and predicted path (and intensity)?

Line 157: "These simulations used a multi-scale ensemble system based on two one-way nested domains to better account for meso- and storm-scale processes involved in the genesis and evolution of Qendresa (Fig. 2)."

Why only one-way nesting is used here (WRF supports two-way nesting if I am not mistaken)?

Line 230: "Following similar studies (e.g., Romine et al. (2013); Yussouf et al. (2015); Carrió and Homar (2016)), the observational error values used here for the conventional observations are: 0.75 K for the temperature, 0.75 K for the dew point temperature, 0.75 m s −1 for the wind speed and 0.75 hPa for the pressure."

In Romine et al. (2013), they use "NCEP statistics" for temperature, Lin and Hubbard (2004) for the dew point temperature, 1.75 m/s for E–W, N–S winds (Buoy and ship reports) and 1 hPa for altimeter (also Buoy and ship reports). Also AMV errors are 50% NCEP statistics in Romine et al. (2013). This seems to me (not a meteorology expert) quite different from the fixed value approach here.

In Yussouf et al. (2015), the used observation errors are described as:

The assumed observation errors are the same as in Table 3 of Romine et al. (2013) except for METAR and marine temperature (1.75 K), METAR altimeter (0.75 hPa), and marine altimeter (1.20 hPa).

I don't understand how the citations are used to support the choice of these parameter values (except for the METAR altimeter) which are crucial for data assimilation (as also noted by the author).

Figure 7 and 9: there are several assimilation runs introduced in the previous section. It is not clear to me which assimilation experiment is presented in Figure 7. Also, I think the author should take more clearly which comparisons are performed against independent data and which use dependent data (used in the analysis). Since for figure 9, a correlation coefficient of 0.996 is achieved for the posterior estimate, I suspect that this is comparison with dependent observation. Much more interesting would be validation with independent data: for example if you assimilated only RSAMV data, do you also improve compared to METARs, rawinsondes and buoys?

Figures: text on Figure 7 and 8 is too small .

Line 370: “In fact, for the SYN experiment only 17/36 ensemble members generate a

small-scale isolated cyclone, while in the RSAMV, a reduced number of members simulate cyclones (16/36), and finally in the CNTRL experiment, this number is increased to 21/36.” How many ensemble members generate a cyclone for the simulation NODA and how many ensemble members in NODA are used for validation later on?

Line 375: “Taking into account this, we have also represented the best track simulated by the different experiments in comparison with the trajectory observed by satellite imagery.” How do you define “best” here?

Line 386: “and the error associated with NODA start to decrease until the end of the simulation.” and Figure 13: Overall the assimilation did not improve a lot compared to the simulation without data assimilation. During 17:00 - 23:00 November 7, the assimilation run has an error which is about 50 km lower, however during the end the error is about 50-100 km larger.

Figure 13: “ensemble spread (shaded areas) of track error (km)”: not all ensemble members present a cyclone. How is this taken into account? If in an ensemble, only one member would have represented a cyclone, would this correspond to a spread of 0 ?

Figure 14: this analysis considers only the center of the medicane; if an ensemble would have a small but consistent biais the medine track, it would result in a very low score when computing the PCC. It would be more useful to compute a probability map for the wind exceeding a given threshold.

Intensity (Figure 15): all experiments underestimate the intensity (indicated by the minimum pressure). But the NODA experiment most realistic mean (in terms of minimum pressure) and a significant fraction of the ensemble members for NODA show the right intensity (but with a time shift) while only few ensemble members with data assimilation for the have the correct intensity (but with a better timing). Timing better and better minimum pressure?

Minor issues:

Line 55: “From these reasons” -> “For these reasons”

Line 104: Add citation for EnKF when mentioning it the first time (after the abstract). There are many variants of the EnKF. The reference is only included in section 3.3.

Line 205: “In other words, one spectral channel can identify the same wind observation that another channel can identify. However, is not common that both different channels provides precisely the same value of such observation.” The formulation is a bit awkward (besides the spelling issues “However, is” -> “However, it is”, “provides” -> “provide”). Please rephrase.

Line 317: rmsi (and legend of plot 7) -> RMSE

Additional references:

Lin, X., and K. G. Hubbard, 2004: Sensor and electronic biases/ errors in air temperature measurements in common weather station networks. J. Atmos. Oceanic Technol., 21, 1025–1032